

*Citation for published version:*

Fincham Haines, T & Wilson, RC 2008, 'Combining shape-from-shading and stereo using Gaussian-Markov random fields', Paper presented at 19th International Conference on Pattern Recognition, ICPR 2008, Tampa, FL, USA United States, 8/12/08 - 11/12/08.

*Publication date:*  
2008

*Document Version*  
Publisher's PDF, also known as Version of record

[Link to publication](#)

**University of Bath**

**Alternative formats**

If you require this document in an alternative format, please contact:  
[openaccess@bath.ac.uk](mailto:openaccess@bath.ac.uk)

**General rights**

Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

**Take down policy**

If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

# Combining Shape-from-Shading and Stereo using Gaussian-Markov Random Fields

Tom S. F. Haines and Richard C. Wilson

Department of Computer Science, University of York, Heslington, York, UK

{thaines,wilson}@cs.york.ac.uk

## Abstract

*In this paper we present a method of combining stereo and shape-from-shading information, taking account of the local reliability of each shape estimate. Local estimates of disparity and orientation are modelled using Gaussian distributions. A Gaussian-Markov random field is used to represent the disparity-map, taking into account interactions between disparity measurements and surface orientation, and the MAP estimate found using belief propagation. Local estimates of the precision of disparities and surface normals are found and used to control the process so that the most accurate data source is used in each region. We assess the performance of our approach using both synthetic and real stereo pairs, and compare against ground truth.*

## 1 Introduction and previous work

Dense stereo algorithms may be divided into two steps. First is the calculation of a matching cost for each disparity at each location, represented by the Disparity Space Image (DSI). In areas with strong cues a DSI gives a clear indication of actual disparity, but in relatively uniform areas it will not distinguish the correct disparity from incorrect disparities. The second step is the selection of disparities to find a consistent solution. Advanced approaches to this problem use techniques such as dynamic programming[1], graph cuts[2] and belief propagation[14]. Shape from Shading (SfS) relies on the shading information available from a single image. It is premised on the intensity of light reflected by a surface being related to the angle between the surface and light source(s)[13]. It therefore provides information about the orientation of the surface. Stereo algorithms do not perform effectively in areas of uniform texture. Such regions will generally either be interpolated or plane fitted, which is not necessarily a true reflection of the surface shape. In contrast SfS can operate only in areas where albedo can be inferred, so a uniform albedo assumption needs to be used. This makes

SfS ideal for filling in areas where stereo has insufficient information[8]. In combining these ideas we have an improved set of modelling assumptions resulting in a surface estimate with greater detail.

The literature on both stereo and SfS is comprehensive and we do not intend to review it in detail here. Here we will focus particularly on methods which combine SfS with stereo. For example, Leclerc and Bobick[8] have used stereo to provide initialisation and boundary constraints for SfS. Cryer, Tsai and Shah[4] combine SfS and stereo in the frequency domain, using a high pass filter for the SfS and a low pass filter for the stereo. Jin, Yezzi and Soatto[7] assume the image is divided into areas of texture and constant albedo and apply separate cost functions to each area and solve with level sets. Shao et al[10] use an additional cost for the difference between SfS irradiance in the left image and image irradiance at the corresponding point in the right image. Motion, stereo and photometric stereo have been integrated into a single framework by Zhang et al[15] and this method can provide accurate object models from a sequence of frames. This can be contrasted with our work where we are interested in recovering shape from two frames only. Also of interest here is the work of Potetz[9] who uses a belief propagation framework to incorporate integrability constraints in the process of surface integration from normals.

## 2 Problem Formulation

Given a rectified colour image pair, left ( $I_L(x, y)$ ) and right ( $I_R(x, y)$ ), we can compute a disparity map,  $D(x, y)$ , representing the correspondences between images. The process may be divided into two steps, first a  $DSI(x, y, d)$  is defined expressing the cost of matching  $I_L(x, y)$  and  $I_R(x + d, y)$ . Modelling assumptions are then used to select an optimal set of matches. Given a camera calibration,  $D(x, y)$  may be converted into world coordinates,  $(X, Y, Z)$ .

SfS uses the (calibrated) grayscale image intensity

of a single image  $L(x, y)$ . The goal of SfS is to recover surface orientation for each pixel,  $\mathbf{n}(x, y)$ . Under a single light source and Lambertian reflectance model the surface normals are related to the image intensity via

$$L(x, y) = A(x, y)\mathbf{n}(x, y) \cdot \mathbf{s} \quad (1)$$

where  $\mathbf{s}$  is the light-source direction and  $A(x, y)$  is the apparent albedo at  $(x, y)$  in the image. The goal of SfS is to recover the surface normals given the luminance map, albedo map and the light source direction. We have the luminance map from the image, and the light source direction is presumed to be known from the camera setup. We need to discover the albedo map  $A(x, y)$  from the images in order to operate SfS. In principle, depth can be recovered from the normal map by integrating over the surface. This is neither straightforward nor accurate however.

### 3 Gaussian-Markov Random Fields

Solving the stereo and SfS problems will provide a field of depths and surface normals respectively, each of varying degrees of accuracy at different points on the surface. Our goal is then to combine these two sources of information to produce an improved estimate of the surface. Belief propagation has been successfully used both for stereo[14] and for surface integration[9] and so we believe it will be a useful approach here. We combine the disparity and orientation information within the framework of Gaussian belief propagation which allows the local probabilistic description of both depth and orientation information. Belief propagation has previously been used with discrete frequency functions to find stereo disparities[11]. In our case we have to recover a continuous disparity otherwise surface normals will not provide much information since the change in disparity implied by the surface normals is often much less than one pixel. One tractable solution is to use Gaussian distributions. The beliefs are then defined continuously by the mean and variance of the Gaussian distribution, allowing orientation information to be used effectively. We adopt this approach in this paper.

In this paper we follow the formulation given by Weiss and Freeman[12]. We need to define two elements; a compatibility distribution between the disparities at  $t$  and  $s$ ,  $\psi_{st}(x_s, x_t)$ , and a distribution of disparities inferred from the observed evidence  $y_t$ ,  $\psi_t(x_t, y_t)$ . Each distribution, and therefore the messages, are represented by a normal distribution. To describe this we adopt a variant of the Gaussian algebra of Cowell[3]. The Normal distribution is defined as a function of the precision  $\mathbf{P}$  and the precision times the mean  $\mathbf{P}\mu$ , which we will refer to as the p-mean. The precision is

equal to the inverse covariance matrix, i.e.  $\mathbf{P} = \Sigma^{-1}$ . We have

$$\phi[\mathbf{P}\mu, \mathbf{P}] = \alpha \exp \left[ -\frac{1}{2}(\mathbf{x} - \mu)^T \mathbf{P}(\mathbf{x} - \mu) \right] \quad (2)$$

The reason for defining  $\phi$  in this way is that it produces a simple set of rules for manipulating the distributions.

A stereo algorithm is used to give an initial estimate of the disparities. At a point  $t$  in the image, the stereo pair gives a set of measurements,  $y_t$ , which are used to infer a distribution for the disparities,  $x_t$ . This is modelled by a Normal distribution,  $\psi_t(x_t, y_t) = \phi[\mathbf{P}_t\mu_t, \mathbf{P}_t]$ . The mean  $\mu$  and precision  $\mathbf{P}$  for this distribution are computed from the stereo algorithm as detailed in section 4.

The compatibility distribution between two neighbouring points in the image  $s$  and  $t$  is also modelled by a Normal distribution. If the disparity at  $t$  is  $x_t$ , then we would expect the disparity at  $s$  to be  $x_t + z_{ts}$  where  $z_{ts}$  is the disparity change predicted by integrating the surface normals along the path from  $t$  to  $s$ . The compatibility distribution  $\psi_{st}(x_s, x_t)$  is therefore defined as a Normal distribution with mean  $x_t + z_{ts}$  and a precision  $P_n$  which reflects the accuracy of the surface normals. We therefore obtain

$$\psi_{st}(x_s, x_t) = \phi \left[ P_n \begin{pmatrix} -z_{ts} \\ z_{ts} \end{pmatrix}, P_n \begin{pmatrix} 1 & -1 \\ -1 & 1 \end{pmatrix} \right] \quad (3)$$

Since the points are neighbours in the image we can assume that the surface normal direction is constant along the path between them, and use an interpolated surface orientation at the half way point. This is in fact necessary to avoid bias in the result. The two separate processes therefore influence the MRF in different ways; the local measurement process models the depth information and the compatibility between sites is used to incorporate the orientation information.

Since the distributions are Normal, the messages are also Normal distributions. The message that  $s$  sends to  $t$  is defined by the distribution parameters  $P_{ts}$  and  $P_{ts}\mu_{ts}$ . We begin by defining the following quantities;  $P_0 = P_t + \sum_{u \in N/s} P_{ut}$  and  $P_0\mu_0 = P_t\mu_t + \sum_{u \in N/s} P_{ut}\mu_{ut}$ . These are the local precision and p-mean respectively, excluding the message we are currently computing. Applying loopy belief propagation, we obtain the update rules:

$$\begin{aligned} P_{ts} &\leftarrow P_n - P_n(P_n + P_0)^{-1}P_n \\ P_{ts}\mu_{ts} &\leftarrow P_n z_{ts} + P_n(P_n + P_0)^{-1}(P_0\mu_0 - P_n z_{ts}) \end{aligned} \quad (4)$$

We iteratively apply these rules to find an estimate of the MAP disparity map. After iteration, the estimated

disparity is given by

$$\mu_t = (P_t \mu_t + \sum_{u \in N} P_{ut} \mu_{ut}) (P_t + \sum_{u \in N} P_{ut})^{-1} \quad (5)$$

#### 4 Local Precisions for Disparity and Orientation

The local confidence in the estimate of disparity can be estimated from the DSI of the stereo pair. The DSI can be considered as a probability distribution for disparities and so firstly, the initial mean  $\mu_t$  is set as the largest weight in the DSI, i.e.  $\mu_t = \arg \max_d \text{DSI}(x, y, d)$ . The variance  $1/P_t$  is then computed from the DSI using a robust m-estimator with Tukey bisquare reweighting[6]. From this the precision and the p-mean,  $P_t \mu_t$ , can be computed. If the variance is above a certain threshold (50 pixels squared) the disparity is considered to be unknown and the precision is set to a vanishing small value ( $10^{-12}$ ). Occluded or background pixels with no disparities are assigned a precision of  $P = 0$ .

For the surface normals, we know that the normals should be accurate in uniform regions and inaccurate in textured areas and region boundaries. We can therefore use a classification based on image gradient to determine areas where normals are likely to be reliable. In areas with colour changes or large intensity gradients, the normals are given a small precision  $P_{n,0}$  and in other regions a larger precision  $P_{n,1}$ . This allows, for example, surface steps at the boundaries of regions, which would not be allowed under the shading model.

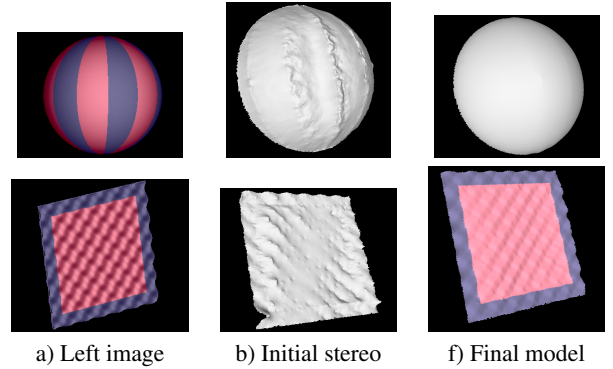
#### 5 Algorithms

We begin with an initial depthmap delivered by a hierarchical belief propagation (HBP) algorithm which is a modified version of the algorithm of Felzenszwalb and Huttenlocher[5]. This provides a rough estimate of the surface normals which we can use to find a reflectance map of the surface. The image is divided into uniform regions using a mean shift algorithm, and reflectance computed on a per-region basis. This process then delivers a shading map on which we apply the SfS method of Worthington and Hancock[13] to obtain surface normals. These constitute the input normal map to the BP algorithm. We then apply our BP algorithm to obtain a refined estimate of the depthmap. We can then in turn compute an improved reflectance map and surface normals. We iterate this process to obtain an accurate object model; three or four iterations are normally sufficient.

### 6 Experimental Results

We have evaluated our method using a number of stereo pairs captured using a stereo camera setup which consists of two parallel mounted cameras. The cameras are standard commercial digital still cameras. The system is fully calibrated geometrically using a calibration target and radiometrically using a light meter. We have obtained ground truth using a Cyberware 3030 3D scanner. This scanner produces a 3D model of the object accurate to within a few millimeters.

We begin with a synthetically generated stereo pair of a textured sphere with added Gaussian noise ( $\sigma = 2$ ). Figure 1 shows the results of this process. In this simple case, the stereo algorithm delivers an accurate surface, but where there is little variation in the image, the surface is too flat. Incorporating the surface shading information improves these areas (Figure 1).



**Figure 1. SSFS results for a synthetically generated data**

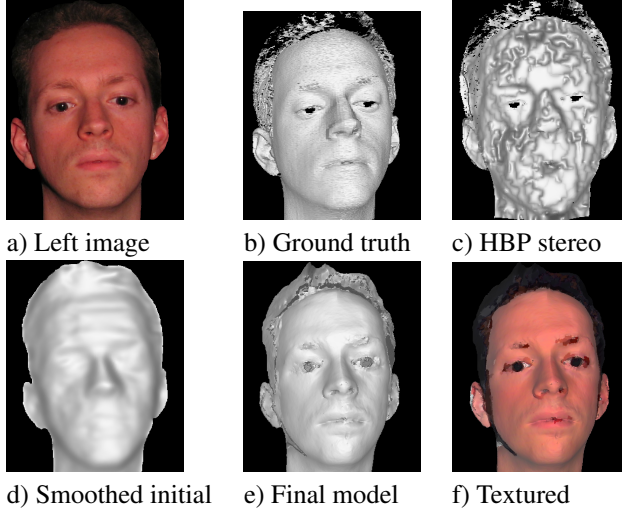
The second stereo pair is of a flat plane which has a small-scale bumpy surface added to it and some limited texture information. Again the images include noise to prevent perfect stereo matching.

	HBP	DP	SfS-BP
Ball	1.05	1.25	<b>0.334</b>
Bumpy plane	0.282	0.335	<b>0.205</b>
Plant pot	45.09	47.11	<b>17.06</b>
Head	5.34	5.39	<b>5.28</b>
Frame	5.94	2.94	<b>1.23</b>

**Table 1. MSE to ground truth disparity.**

	HBP	DP	SfS-BP
Ball	0.0366	0.0414	<b>0.0199</b>
Bumpy plane	0.1255	<b>0.0421</b>	0.0574
Plant pot	0.384	0.494	<b>0.0771</b>
Head	0.208	0.258	<b>0.114</b>
Frame	0.215	0.339	<b>0.0665</b>

**Table 2. Error in the surface normals.**  
We now turn our attention to image pairs captured



**Figure 2. The SfS-BP process for the head**

by our stereo camera, one of which is shown in Figure 2. This image shows a large improvement in the fine surface detail of the models.

Table 1 shows the mean square error between the recovered disparity and ground truth disparity for a number of datasets and different stereo methods. The methods used are hierarchical belief propagation(HBP)[5], Dynamic programming(DP)[1], and our method (SfS-BP). In all cases there is an improvement in the reconstructed disparity map from using the shading information through our belief propagation. The picture frame and plant pot both show large improvements. The picture frame has raised relief which is not accurately picked up by the stereo algorithms but is modelled via the shading. In table 2 we analyse the error in the surface normals computed from the disparity map (not from SfS). The error in the surface normals is measured by  $1 - \mathbf{n} \cdot \mathbf{n}_t$  where  $\mathbf{n}$  is the measured surface normal and  $\mathbf{n}_t$  is the ground truth surface normal, so we obtain 0 if the normals are identical, and 1 if they are randomly distributed. The big advantage of incorporating shading information is in the accuracy of the surface normals, as shown in Table 2 where there is a large improvement for all but the plane model. This leads to much more realistic-looking models.

## 7 Conclusions

We have presented a method of integrating shape from shading information with stereo information using Gaussian belief propagation. This method efficiently delivers a continuous estimate of disparity and allows separate control of the local confidence in the disparity and normal information, via appropriately defined Gaussian distributions. We can therefore use stereo in-

formation where it is reliable, and shading information in other areas. We have presented a number of experiments with our method, both on synthetic and real image pairs. Comparison with ground truth shows that this method is effective in improving disparities and produces large improvements in surface normal information and therefore model realism.

## References

- [1] A. A. Amini, T. E. Weymouth, and R. C. Jain. Using dynamic programming for solving variational problems in vision. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 12(9):955–867, 1990.
- [2] Y. Boykov, O. Veksler, and R. Zabih. Fast approximate energy minimization via graph cuts. *IEEE transactions on Pattern Analysis and Machine Intelligence*, 23(11):1222–1239, 2001.
- [3] R. Cowell. *Learning in Graphical Models*, chapter Advanced Inference in Bayesian Networks. MIT Press, 1998.
- [4] J. E. Cryer, P. S. Tsai, and M. Shah. Integration of shape from shading and stereo. *Pattern recognition*, 28(7):1033–1043, 1995.
- [5] P. F. Felzenszwalb and D. P. Huttenlocher. Efficient belief propagation for early vision. *IEEE CVPR*, 1:261–268, 2004.
- [6] P. J. Huber. *Robust Statistics*. Wiley, 1981.
- [7] H. Jin, A. Yezzi, and S. Soatto. Stereoscopic shading: Integrating multiframe shape cues in a variational framework. In *CVPR*, volume 1, pages 169–176, 2000.
- [8] Y. G. Leclerc and A. F. Bobick. The direct computation of height from shading. In *CVPR*, pages 552–558, 1991.
- [9] B. Potetz. Efficient belief propagation for vision using linear constraint nodes. In *CVPR*, 2007.
- [10] M. Shao, R. Chellappa, and T. Simchony. Reconstructing a 3-d depth map from one or more images. *CVGIP: Image Understanding*, 53(2):219–226, 1991.
- [11] J. Sun, N.-N. Zheng, and H.-Y. Shum. Stereo matching using belief propagation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 25(7):787–800, 2003.
- [12] Y. Weiss and W. T. Freeman. Correctness of belief propagation in gaussian graphical models of arbitrary topology. *Neural Computation*, 13(10):2173–2200, 2001.
- [13] P. Worthington and E. Hancock. New constraints on data-closeness and needle map consistency for shape-from-shading. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 21(12):1250–1267, 1999.
- [14] J. Yedidia, W. Freeman, and Y. Weiss. Constructing free-energy approximations and generalized belief propagation algorithms. *IEEE Transactions on Information Theory*, 51(7):2282–2312, 2005.
- [15] L. Zhang, B. Curless, A. Hertzmann, and S. M. Seitz. Shape and motion under varying illumination: Unifying structure from motion, photometric stereo, and multi-view stereo. In *ICCV*, pages 618–625, 2003.